PASSIVE SMOKING AND LUNG CANCER: REANALYSES OF HIRAYAMAS DATA

W.Ahlborn[1] & K.Überla[2]

[1]Institut für Statistik und Ökonometie der Universität Göttingen Platz der Göttinger Sieben 5,D 3400 Göttingen, FRG

[2]Institut für Med. Informationsverarbeitung, Statistik und Biomathematik, Marchioninistraße 15 D, 8000 München 70, FRG

ABSTRACT

The cohort investigated by HIRAYAMA has a serve selection bias by age. The effect of removing this selection bias by iterative proportional fitting a contingency table to given marginals is investigated. The risk increase reported by HIRAYAMA disappears completely when one removes selection bias by age. If the cases would have been observed as they occur in the female population one would have observed no risk increase. Only in the subgroup of women married to industry workers there remains a risk increase, which might be due to confounding factors. Assuming modest differential misclassification also leads to risk ratios around unity.

INTRODUCTION

The statistical association between enviromental tobacco smoke and lung cancer is controversial. The HIRAYAMA study seems to provide sound epidemiological evidence supporting this hypothesis. In a recent paper ÜBERLA (6) has analysed the published studies. Regarding the HIRAYAMA study the following facts have to be kept in mind:

- The study was not designed to test the hypothesis, whether passive smoking is associated with lung cancer or not. It can therefore only generate this hypothesis, not prove it.
- The cohort was not representative for the population of Japan. A selection bias is possible.
- The exposure indicator - the fact of being married to a man who smokes - is not reliable, not valid and not specific.
- The event indicator - dying on lung cancer as noted on death certificates - is neither reliable nor valid.
- Various confounding factors - for instance exposure at the working place, indoor air pollution, overall air pollution, type of medical care - were not accounted for.
- Bias in registering the fact, that a woman is a nonsmoker, was not controlled. Resulting differential misclassifications of the cases, who were smokers and had to be excluded, have not been considered.
- Almost nothing is known about the 200 cases. No case reports are available, autopsy and histology are only available in 11.5 %.

The core of the information, on which the results of this study rely, is

1.) that during 1965 200 women in Japan told an interviewer on a single occasion that they were - during that time - non-smokers and their husbands told, that they were smokers, which might have been different before and afterwards and
2.) that their death certificates subsequently contained the diagnosis lung cancer, which might have been erroneous.

Such sparse information does not seem to be convincing.

In our paper we consider four questions:
1.) What is the relative risk when one removes the selection bias regarding age of women in the HIRAYAMA cohort?
2.) What is the relative risk when one additionally accounts for the fact that women above 70 who are married to busbands still living are less frequent than reported in the population statistic?
3.) What is the relative risk for women married to men with different occupations, when one removes the selection bias regarding age of men?
4.) What is the relative risk when additionally some modest differential misclassification is assumed?

MATERIALS AND METHODS

We start from tables 1, 2 and 3 of HIRAYAMA 1984 (4). These tables contain the most detailed published data. In order to check our program we reproduced some of the reported relative risks with good accuracy.

PERCENT FEMALE

AGE GROUP	JAPAN POPULATION	HIRAYAMA COHORT
40-49	39	42
50-59	30	35 ↑
60-69	19	22
70 +	12	1 ↓
	100	100

TABLE 1: Differences between the HIRAYAMA cohort and the female age distribution over 40 in the population of JAPAN 1965 (Population census 1965. Statistical survey of the economy of Japan. 1967, Ministry of Foreign Affairs of Japan).

There are marked differences between the HIRAYAMA cohort and the female age distribution over 40 in the population of Japan 1965. Women 50-59 are over-represented, women older than 70 are severely underrepresented. In this age group only one percent was observed instead of 12 percent in the population. The investigated cohort certainly has a severe selection bias by age, which needs no statistical test. This is likely due to the fact, that the smoking behaviour was not known in the elderly or that the husbands of older women have died. Since it takes twenty years and more from exposure to lung cancer, older women surely are relevant and should not be excluded. The majority of lung cancer cases occur in older age groups, in Germany more than 67 % in women over 65 years.

In order to answer the question what the relative risk is when the age selection bias is removed, we adjusted the data to the age distribution of the female population of Japan. The technique of iterative proportional fitting a contingency table to given marginals as described by BISHOP, FIENBERG and HOLLAND (1) or by HARTUNG (3) was used. This technique keeps the risks constant as observed in every cell and changes the marginals and the cell counts according to the given age distribution of the population. Iterative proportional fitting of contingency tables to given marginals is a well known technique in multivariate statistics and can be applied here without changing the observed interrelations between smoking habit, occupation and lung cancer. From the fitted or adjusted tables the risk ratios are calculated in the usual way. Such risk ratios based on data with removed age selection bias are the correct ones and should be used. One has to require that there should be no selection bias by age and the cases should be included as they would have occured in the population. Otherwise statistical tests and P-values are not very meaningful.

WIVES AGE	HUSBAND'S SMOKING HABITS NON		1 - 19		20 +		TOTAL	
40-49	4	7918	21	17492	21	12615	46	38025
50-59	14	7635	46	15640	31	8814	91	32089
60-69	16	6170	31	10381	10	3793	57	20344
70 +	3	172	1	671	2	239	6	1082
TOTAL	37	<u>21895</u>	99	<u>44184</u>	64	<u>25461</u>	200	<u>91540</u>

TABLE 2: SMOKING HABIT OF HUSBAND BY AGE OF WIFE. ORIGINIAL DATA (Table 2 of HIRAYAMA 1984).

Table 2 shows the original data by age of wife. The cells contain the number of lung cancer cases and those under risk as published by HIRAYAMA. The 1-19 group includes ex-smokers in this and the following tables. 200 cases out of 91540 women were observed. Iterative proportional fitting to the female age distribution of the population leaves the underlined numbers constant. The others are adjusted using a rigth hand marginal which is made proportional to the age distribution of the population.

RESULTS

Table 3 gives the results of iterative proportional fitting to the female age distribution of the population. It contains the numbers of those under risk and of lung cancer deaths as they would have been ovserved, if HIRAYAMA had not excluded or preferred certain age groups. The age selection bias is removed. The risks in the individual cells are still the same as those observed by HIRAYAMA. Also the structure of the common distribution regarding age, smoking habit and lung cancer is unchanged. HIRAYAMA would have totally observed 232 cases instead of 200, with the corresponding numbers in the individual cells, had he included all women as they live in the population. This table is the best available starting point for age-adjusted risk ratio calculations. It was not used so far.

WIVES AGE	NON		HUSBAND'S SMOKING HABITS 1 - 19		20 +		TOTAL	
40-49	3.91	7784.8	19.12	15927.8	20.02	12024.0	43.05	35700.6
50-59	12.49	6813.7	38.20	12987.1	26.95	7661.2	77.64	27462.0
60-69	14.25	5496.6	25.70	8604.9	8.68	3291.1	48.63	17392.6
70 +	32.02	1835.2	9.93	6664.2	20.79	2484.7	62.74	10984.8
TOTAL	62.67	21895	92.95	44184	76.44	25461	232.06	91540

TABLE 3: SMOKING HABIT OF HUSBAND BY AGE OF WIFE (Table 2 of HIRAYAMA 1984). Removed selection bias: Data adjusted to the age distribution of women in the population.

	HUSBAND'S SMOKING HABITS NON	1-19	20 +
$\hat{RR}$	1.00	1.37	1.56
IL_{90}		1.00	1.11
MH-CHI		1.51	2.27
$P_{\text{ONE TAILED}}$		.065	.012*
$\hat{RR}$	1.00	.77	1.06
IL_{90}		.59	.80
MH-CHI		2.19	.27
$P_{\text{ONE TAILED}}$		.014**	.395

UPPER PART : STANDARDIZED BY AGE OF WOMEN ONLY
LOWER PART : AGE SELECTION BIAS REMOVED AND STANDARDIZED BY AGE OF WOMEN

RR : Weighted point estimate of rate ratio
IL_{90} : Lower 90 percent confidence interval
* : "significant" in positive direction
** : "significant" in negative direction

TABLE 4: RELATIVE RISK BY AGE OF WOMEN (Calculated from table 2 of HIRAYAMA 1984)

In the upper part of table 4 one finds the risk ratios standardized by age only, as reported by HIRAYAMA. The lower part contains the risk ratios after removing the age selection bias. In the upper part the weighted point estimate of the rate ratio is 1.56 in the 20 + group and is technically "significant". IL_{90} designates the lower point of the 90-percent confidence interval in this and the following tables, as it was used by HIRAYAMA.

The risk increase disappears completely when one removes the selection bias by age. In the 20 + group the rate ratio is 1.06, hardly a relevant risk increase. In the group of 1-19 cigarettes per day it is .77 which is a technically significant risk decrease. The adjusted rate ratio, considering all those exposed in one group versus those not exposed is .81 with a confidence interval including unity. If HIRAYAMA had observed the cases as they occur in the population without selection bias by age, he would have observed no risk increase, but a sligth and meaningless risk decrease. This is the main result of our reanalysis, which corresponds well with the result of the prospective American cohort study as published by GARFINKEL (2).

In the discussion following our paper in TOKIO last november HIRAYAMA noted, that in the population the percentage of women over 70 married to men who are still alive is smaller than the percentage of women reported in the population statistics. Since we do not have the numbers we assume that only half of the women over 70 reported in the population census 1965 have been married to living husbands. The resulting hypothetical population together with the HIRAYAMA cohort is presented in table 5.

AGE GROUP	PERCENT FEMALE HYPOTHETIC POPULATION	HIRAYAMA COHORT
40-49	42	42 ↑
50-59	32	35
60-69	20	22
70 +	6	1 ↓
	100	100

TABLE 5: DIFFERENCES BETWEEN THE HIRAYAMA COHORT AND A HYPOTHETIC FEMALE AGE DISTRIBUTION OVER 40. (Explanation see text)

There is still possibly a selection bias in table 5. Now 6 percent of women over 70 would have been included in the hypothetic female distribution instead of 12 percent. The corresponding lung cancer cass, which generally are more frequent in this age group than in younger women, had been excluded. The reduction to one half accounts for the argument of HIRAYAMA mentioned above sufficiently. The resulting relative risks are presented in table 6. Even with these assumptions the relative risk is only 1.03 in the group of women married to husbands smoking 1-19 cigarettes per day, 1.29 in the 20 + group and 1.12 if one considers the smoking group altogether. All these risk ratios are not statistically different from unity.

	HUSBAND'S SMOKING HABITS			
	NON	1-19	20 +	SMOKER
$\hat{RR}$	1.00	1.03	1.29	1.12
IL_{90}		.77	.94	.85
MH-CHI		.05	1.33	.47

TABLE 6: RELATIVE RISK BY AGE OF WOMEN, AGE SELECTION BIAS REMOVED AND HYPOTHETICALLY ADJUSTED TO HIRAYAMAS ARGUMENT (see table 5)

2023513496

Since it is impossible for us to reconstruct the real situation some twenty years ago in Japan regarding the conditional distributions of males and females regarding age, smoking and family status, the reported results of the HIRAYAMA study can not be conclusive to us. As long as the selection bias by age can not be explained numerically in a sufficient way by HIRAYAMA, his thesis, that there is a significant and relevant risk increase based on his data migth as well be wrong.

We now consider two occupations, farmers and industry workers. From the upper part of table 7 one can see that the relative risk for wives of farmers seems substantial, when one standardizes by age of men only. The point estimates of the rate ratios are 1.48 and 1.63 respectively. This was observed earlier and had no adequate exlanation. If one removes the selection bias by age and adjusts to the male age distribution of Japan - the numbers in the lower part of table 7 - the rate ratios are .85 and .82, not different from unity. This seems more plausible.

	HUSBAND'S SMOKING HABITS NON	1-19	20 +
$\hat{RR}$	1.00	1.48	1.63
IL_{90}		.97	1.01
MH-CHI		1.48	1.92
$P_{ONE\ TAILED}$		.069	.027
$\hat{RR}$	1.00	.85	.82
IL_{90}		.59	.53
MH-CHI		.42	.53
$P_{ONE\ TAILED}$		.337	.296

UPPER PART : STANDARDIZED BY AGE OF MEN ONLY
LOWER PART : AGE SELECTION BIAS REMOVED AND STANDARDIZED BY AGE OF MEN

TABLE 7: RELATIVE RISKS: WIVES OF FARMERS ONLY
(Table 3 of HIRAYAMA 1984)

Considering the wives of industry workers only, in the upper part of table 8 the point estimates of the rate ratios are 1.77 and 2.27, standardized by age of men, being not significant. Removing the age selection bias - in the lower part of table 8 - there is a remarkable risk increase to 4.60 and 6.90, which is significant. However, there are only 9 lung cancer deaths in the 20+ group and only 3 in women 70 years and older, which are small numbers, but these are numbers observed and used by HIRAYAMA and his risk structure is unchanged. Thus only in the subgroup of women married to industry workers there is a risk increase, in all other occupations there is no risk increase. Omitting industry workers, the point estimates of the rate ratios are .90 and .89, not significantly different from unity. These findings are consistent with the assumption of confounding factors in women married to industry workers, who migth be exposed to other enviromental hazards. Our calculations show that by removing age selection bias by age, one can explain hitherto implausible results.

	HUSBAND'S SMOKING HABITS NON	1-19	20 +
$\hat{RR}$	1.00	1.77	2.27
IL_{90}		.70	.84
MH-CHI		.73	.81
$P_{ONE\ TAILED}$		.232	.208
$\hat{RR}$	1.00	4.60	6.90
IL_{90}		1.71	2.45
MH-CHI		2.50	2.78
$P_{ONE\ TAILED}$		.006	.003

UPPER PART : STANDARDIZED BY AGE OF MEN ONLY
LOWER PART : AGE SELECTION BIAS REMOVED AND STANDARDIZED BY AGE OF MEN

TABLE 8: RELATIVE RISKS: WIVES OF INDUSTRY WORKERS ONLY (Table 3 of HIRAYAMA 1984)

Active smoking is correlated among married couples. In a society in which female smokers were very rare in 1965, more women married to smokers will declare themselves nonsmokers than the other way round. One has therefore to consider biased or differential misclassification. There are likely more women with lung cancer, who have been misclassified as nonsmokers than the other way round. They have to be removed from the cohort. We made some moderate assumptions regarding misclassification, as shown in table 9. In order to examine, how sensitive the relative risk is we removed 10, 20 and 30 cases from the exposed groups corresponding to 5, 10 and 15 percent. Assuming 30 misclassified cases - 15 percent, a percentage which has been observed in the literature (5) - the rate ratios are .66 and .85. In the group 1-19 cigarettes per day all the risk estimators are significantly smaller than unity. Our personal opinion is that 10 differential misclassified cases from 200 is a fair number. The corresponding weigthed point estimates of the rate ratio are .74 and 1.00. These risk estimates are as reasonable as other risk estimates calculated from the HIRAYAMA data. They indicate - if anything - a risk decrease, not a risk increase.

NUMBER OF CASES ASSUMED MISCLASSIFIED AND REMOVED FROM EXPOSED GROUPS		HUSBAND'S SMOKING HABITS NON	1-19	20 +
n = 10 = 5 %	$\hat{RR}$	1.00	.74	1.00
	$P_{ONE\ TAILED}$		.006	.469
n = 20 = 10 %	$\hat{RR}$	1.00	.70	.93
	$P_{ONE\ TAILED}$		.003	.383
n = 30 = 15 %	$\hat{RR}$	1.00	.66	.85
	$P_{ONE\ TAILED}$		.001	.238

TABLE 9: RELATIVE RISK: ASSUMED DIFFERENTIAL MISCLASSIFICATION
(Age selection bias removed and standardized by age of women)

DISCUSSION

Reanalyses of data, which have been collected by others are not easy. This is because information is not completely available, because information might be misinterpreted or because one has to take another view in order to come closer to the acceptable truth. Our calculations do not diminish the great value and impact the HIRAYAMA study had on the epidemiology of passive smoking. They show however, that reasonable alternative views on the same data are possible, which lead to opposite conclusions. Our findings are in contrast to HIRAYAMA's thesis that - based on his data - there is a substantiall statistical association between passive smoking and lung cancer. We do not hold that our view is the only correct one. We do hold however, that the risk ratios calculated by us, removing age selection bias, are as reasonable as the ones calculated by HIRAYAMA. Since they go back to the population and not to a selected sample our estimates could be preferable. Hypothetically accounting for the argument of HIRAYAMA, that in the population the percentage of women over 70 married to men who are still alive is smaller than the percentage of women reported in the population statistics does not change our results. Our risk estimates are a consequence of the data published by HIRAYAMA and can not be rejected from the study data, as they are published so far.

2023513499

		HUSBAND'S SMOKING HABITS NON	1-19	20 +
AGE SELECTION BIAS REMOVED AND AGE-STANDARDIZED (WOMEN)	$\hat{RR}$	1.00	.77	1.06
	$P_{ONE\ TAILED}$		.014	.395
WITHOUT INDUSTRY WORKERS, AGE SELECTION BIAS REMOVED AND AGE-STANDARDIZED (MEN)	$\hat{RR}$	1.00	.90	.89
	$P_{ONE\ TAILED}$		.394	.179
10 CASES ASSUMED MISCLASSIFIED, AGE SELECTION BIAS REMOVED AND AGE-STANDARDIZED (WOMEN)	$\hat{RR}$	1.00	.74	1.00
	$P_{ONE\ TAILED}$		.006	.469

TABLE 10: REANALYSIS OF HIRAYAMAS DATA: SUMMARY OF RELATIVE RISKS

To summarize: Removing the age selection bias in the HIRAYAMA study one gets a relative risk of 1.06 in the group of women married to men with more than 20 cigarettes per day. In the group of women married to men with 1-19 cigarettes per day the relative risk is .77, a technically "significant" risk decrease. If HIRAYAMA could have observed the lung cancer cases as they occur in the female population, he would have observed no risk increase, but a risk decrease to around .81, not significantly different from unity, considering those exposed versus those not exposed.

If one omits the wives married to industry workers because of possible confounding factors the relative risk is .90 and .89 respectively. This is of the same size order and smaller than unity. Here we could adjust and standardize by occupation and age of men only, which is not as appropriate as by age of women.

If one assumes that 10 cases are differentially misclassified and removes them from the exposed groups, the risk estimates are .74 and 1.00 respectively.

Our findings demonstrate how sensitive the data of this study are and how week the evidence for a statistical association between passive smoking and lung cancer from this study is. In view of these and other facts, which we mentioned in the introduction, the null hypothesis might be true as well and is consistent with the HIRAYAMA data in the same way as the alternative hypothesis.

We would be glad to apply our technique to more detailed data if we can get them from HIRAYAMA, for instance in order to adjust by occupation of men and age of women, or by occupation of men and by age of women married to a husband who is still alive. We are ready to modify our view if such data can support the alternative hypothesis better than the published data do. We do hope, that our calculations give rise to a fruitful discussion. The methods we used here might be of interest to the analysis of other cohort and control studies.

2023513500

References

1. Bishop YV, Frienberg StE, Holland PW
Discrete Multivariate Analysis: Theory and Practice
MIT Press, Cambridge p 97 (1980)

2. Garfinkel L, Time trends in lung cancer mortality among nonsmokers and a note on passive smoking, J Natl Canc Inst 66 : 1061-1066 (1984)

3. Hartung J, Elpelt B, Klösener KH
Statistik. Lehr- und Handbuch der angewandten Statistik.
München - Wien - Oldenburg, pp 501-503 (1985)

4. Hirayama T, Lung cancer in Japan: effects of nutrition and passive smoking. In: Mizell M, Correa P (eds.) Lung cancer: Causes and prevention. Verlag Chemie, Weinheim, pp 175-195 (1984)

5. Lee PN, Chamberlain J, Alderson HR, Relationship of passive smoking to risk of lung cancer and other smoking-associated diseases. Br. J. Cancer 54: 97-105 (1986)

6. Überla K, Lung cancer from passive smoking: hypothesis or convincing evidence? Int Arch Occup Environ Health 59: 421-431 (1987)